REMARKS

Applicants have amended claims 1, 9, and 17 and have added new claims 25-27. In view of the above amendments and the following remarks, reconsideration of the outstanding office action is respectfully requested.

The Office has rejected claims 1, 2, 6, 8, 7, 9, 10, 14, 15, 17, 18, and 22-24 under 35 U.S.C. 103(a) as being unpatentable over U.S. Patent No. 581,666 to Alshawi (Alshawi) in view of US 5,818,441 to Throckmorton et al. (Throckmorton), claims 3, 11, and 19 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton et al, and further in view of US. Patent: 5,929,927 to Rumreich et al (Rumreich), Claims 5, 13, and 21 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton et al. and further in view of US. Patent No. 6,513,003 to Angell et al (Angell), and Claim 16 under 35 U.S.C. 103(a) as being unpatentable over Alshawi in view of Throckmorton et al, and further in view of U.S. Patent No. 5,900,908 to Kirkland et al (Kirkland).

The Office asserts Alshawi discloses at least one processor for providing real-time subtitles [captioning] in an AV signal which includes the automatic conversion of an audio [including speech] signal in the AV signal. The Office asserts that Alshawi describes in Fig 1., a video-based communications device (5,8) that provides segmentation of an AV signal (16) and the further processing of the audio [speech] portion of the signal to provide continuous speech-to-subtitles [speech-to-text] translation (19,21,22) that has the ability to overlay and display text subtitles onto AV signal in real-time [captioning] (26). The Office asserts that Alshawi does not disclose synchronizing the caption data with one or more cues in the AV signal, but asserts that Throckmorton et al. teaches a data synchronizing sub-system whose function is to synchronize the primary data stream generated by sub-system 10 with specific associated data. The Office asserts that input to data synchronizing sub-system 20 is scene information from the primary data stream in the form of timecodes and time durations [cues] and data from associated data generator sub-system 16. The Office asserts it creates a so called script for the delivery and display of associates(data at specific points in time. The Office asserts the ability to synchronize the associated data with the primary has many benefits including helping the hearing impaired viewers to better understand AV content and providing relevant data such as synchronized captions that pertain to a television broadcast in real-time (Throckmorton, providing real-time associated data, Col. I, Lines 59-

67). The Office asserts that the combination of Alshawi and Throckmorton et al. discloses an integrated method and apparatus for providing real-time subtitles (captioning) in an AV signal, but does not specifically suggest a method of converting the audio portion of the signal to text data that checks whether the amount of caption data is great than a threshold amount, however the Office asserts that Rumreich teaches the processing of additional captioning data if a counter indicates that a caption buffer threshold has been exceeded (Col. 10, Lines 16-34). The Office also asserts that Alshawi does not show the embedding [encoding] of the text *[caption]* data within the AV signal, but asserts that Angell teaches the embedding [encoding] of the text [caption] data within the AV signal (Fig. 1(108, 140); Col 4, Line 55 - Col 5, Line 17). The Office asserts that the combination of Alshawi and Throckmorton do not show portability and the utilization of the device in the classroom, but asserts that Kirkland teaches a method of providing encoding caption data into the program signal.

Alshawi, Throckmorton, Rumreich, Angell, and Kirkland, alone or in combination, do not disclose or suggest, "selecting a number of lines of caption data which can be displayed at one time . . . automatically identifying a voice and speech pattern in an audio signal from a plurality of voice and speech patterns with the speech-to-text processing system . . . training a speech-to-text processing system to learn one or more new words in the audio signal . . . wherein the direct translation is adjusted by the speech-to-text processing system based on the training and the identification of the voice and speech pattern . . . associating the caption data with the AV signal at a time substantially corresponding with the converted audio signal in the AV signal from which the caption data was directly translated with the speech-to-text processing system. . . displaying the AV signal with the caption data at the time substantially corresponding with the converted audio signal in the AV signal, wherein the number of lines of caption data which is displayed is based on the selection." as recited in claims 1 and 17 or "a speech-to-text processing system that selects a number of lines of caption data which can be displayed at one time, identifies a voice and speech pattern in an audio signal, trains to learn one or more new words in the audio signal, and directly translates ~~converts~~ an audio signal in an AV signal to caption data based on the training and the identification of the voice and speech pattern . . . a signal combination processing system that associates the caption data with the AV signal at a time substantially corresponding to the converted audio signal in the AV signal from which the caption data was directly translated with the speech-to-text processing system, wherein the signal combination

processing system synchronizes the caption data with one or more cues in the AV signal . . . a display system that displays the AV signal with the caption data at the time substantially corresponding with the converted audio signal in the AV signal, wherein the number of caption data which is displayed is based on the selection" as recited in claim 9.

Alshawi discloses in FIG. 1 and at col. 2, line 47 to col. 3, line 28, that an audio signal 17 is fed into a recognizer 19 which is a conventional speech recognizer which converts human speech to text. Alshawi uses a recognition hypothesis 20 which consists of one or more possible sequences of words in text format that corresponds to the audio signal and provides a signal representing the most likely textual counterpart to the spoken language. Alshawi also discloses a translator that translates the signal into a target language and a subtitle generator 24 which overlays the text and video signal to create a display signal. However, nowhere does Alshawi disclose or suggest selecting a number of lines of text which can be displayed at one time and then displaying based on the selection. Additionally, there is no teaching or suggestion in Alshawi of automatically identifying the particular voice and speech pattern in the audio signal from a plurality of voice and speech patterns or in training the recognizer to learn one or more new words in the audio signal. As a result, Alshawi also does not teach adjusting the translation by the recognizer based on any training and on the identification of the particular voice and speech pattern. Rumreich only discloses at col. 4, line 63 to col. 4, line 12, that two lines of text are displayed and that as the amount of text increases the rate of scrolling increases. However, Rumreich does not disclose or suggest selecting a number of lines of text which can be displayed at one time and then displaying based on the selection or any of the other recited limitations. Angell only discloses at col. 4, lines 14-46, that the editorial agent reviews the raw textual data stream produced by the voice recognition server to correct mistakes, not to learn one or more new words in the audio signal or to adjusting the automatic translation based on this learning. Angell also does not disclose or suggest the other recited limitations. Like Alshawi, Rumreich and Angell, the other cited references also do not disclose or suggest the claimed invention.

The present invention provides a method and system with a number of features that provide for a synchronized, easily readable, and accurate captioning of an AV signal. The present invention enhances the viewing and learning experience of the hearing impaired because it eliminates many of the prior distractions which occurred with prior technologies.

With the present invention, the quantity of captioning data which is displayed at one time can be selected to be at an amount that can be digested by the hearing impaired viewer. For example, this amount may vary based on the particular education level of the audience. Accordingly, classroom presentations for grade schools might contain fewer lines of caption data then a college graduate level course. Additionally, with the present invention the captioning data is more accurate than prior systems because the present invention learns new words and can identify a voice and speech pattern from a plurality of voice and speech patterns which assists with the accuracy of the translation process. Further, with the present invention the captioning data is synchronized with the initial audio signal so that the words coming out of the speaker's mouth on the audio-video signal correspond with the caption data being displayed. As a result, a hearing impaired individual can read the caption data and/or lip read from the speaker's mouth because the caption data and the audio signal have been synchronized. Accordingly, the combination of these enhancements greatly improves the ability of a hearing impaired individual to comprehend and retain the content of the audio-video signal and to enjoy the viewing experience.

Accordingly, in view of the foregoing amendments and remarks, the Office is respectfully requested to reconsider and withdraw the rejection of claims 1, 9, and 17. Since claims 2-3 and 5-8 depend from and contain the limitations of claim 1, claims 10-11 and 13-16 depend from and contain the limitations of claim 9, and claims 18-19 and 21-24 depend from and contain the limitations of claim they are patentable in the same manner as claims 1, 9, and 17.

Applicants have also added new dependent claims 25-27. None of the cited references, alone or in combination, disclose or suggest, determining the type of the encoder being used with the speech-to-text processing system and then retrieving settings for the speech-to-text processing system to communicate with the encoder based on the identification of the encoder. As set forth on page 11, lines 22-28 in the above-identified patent application, "Additionally, since various types of encoders 40 may be utilized with speech-to-text processing system 20, a user may cause the AV captioning system 10 to perform a handshake process to recognize a newly added encoder 40 by clicking on initializing encoder button 88, which causes the speech-to-text processing system 20 to recognize which type of encoder 40 is being used and to retrieve its settings so the speech-to-

text processing system 20 can communicate with the device." Accordingly, these dependent claims are also believed to be distinguishable over the cited references and in condition for allowance. A notice to that effect is respectfully requested.

In view of all of the foregoing, applicant submits that this case is in condition for allowance and such allowance is earnestly solicited.

Respectfully submitted,

Date: Sept 20, 2005

Gunnar G. Leinberg
Registration No. 35,584

NIXON PEABODY LLP
Clinton Square, P.O. Box 31051
Rochester, New York 14603-1051
Telephone: (585) 263-1014
Facsimile: (585) 263-1600

CERTIFICATE OF MAILING OR TRANSMISSION [37 CFR 1.8(a)]

I hereby certify that this correspondence is being:

☒ deposited with the United States Postal Service on the date shown below with sufficient postage as first class mail in an envelope addressed to: Mail Stop Amendment, Commissioner for Patents, P. O. Box 1450, Alexandria, VA 22313-1450

☐ transmitted by facsimile on the date shown below to the United States Patent and Trademark Office at (703) _____.

9/21/05
Date

Signature

Sherri A. Moscato
Type or Print Name